

Análisis de satisfacción de las madres solteras en Colombia

Asignatura

Técnicas de aprendizaje estadístico

Universidad Nacional de Colombia

Sede Medellín

Facultad de minas

2018-2

Profesor:

Juan David Ospina Arango

Hecho por:

Santiago Restrepo Álvarez

Mateo Zuluaga Giraldo

Sebastian Ramírez Tamayo

David Ossa Saldarriaga

Daniel Restrepo Mejía

Introducción:

El siguiente trabajo es un análisis hecho con base en una encuesta hecha por el DANE, llamada “Encuesta Nacional de Calidad de Vida” sobre la que se profundizará más adelante. Los datos y tablas provenientes de esta encuesta, están disponibles para que cualquiera los pueda descargar (los datos están anonimizados) y realizar el análisis que quiera. Para nuestro trabajo, se nos encomendó analizar las variables y cuestiones que afectan la satisfacción de las madres solteras en Colombia, y para esto es necesario un breve entendimiento de lo que se interpretará como madre soltera en el trabajo.

Entendimiento inicial:

Se denomina “Madre soltera” generalmente a un tipo de familia monoparental, en la que una mujer lleva a cabo la crianza de los hijos y el manejo del hogar sin la compañía o apoyo de una pareja, por decisión propia o circunstancias de su entorno. Puede referirse también a la madre que no se encuentra unida o casada con su pareja. Para este análisis se toma como madre soltera a las mujeres que sean solteras, viudas o divorciadas y que además tengan algún hijo que viva en el mismo hogar, más adelante se explicará más a fondo el proceso de filtrado.

Datos técnicos de la encuesta:

Para el análisis hecho, se tomó la encuesta hecha en el año 2017, sin embargo, en la página del DANE se puede obtener las encuestas hechas en los siguientes años: 1993, 1997, 2003, 2007, 2008 y de 2010 a 2017. En los primeros años que se hizo la encuesta, esta no se hacía sobre una muestra representativa de todo el territorio nacional, sino que se hacía solo en ciertas ciudades o departamentos, además la encuesta inicialmente no se realizaba anualmente, pero después de los resultados obtenidos en 2010, se tomó la decisión de hacerla con la periodicidad de un año y se añadieron más preguntas con respecto a las encuestas anteriores posteriores, se hicieron cambios estructurales en la encuesta, llegando a la del año 2017 que sería la más actual y fiel para realizar el análisis. Tipo de datos: Encuesta por muestreo. Unidades de análisis: corresponden a las viviendas, los hogares y las personas. El alcance de la encuesta está dado por las siguientes temáticas o grupos de preguntas:

Tipo de vivienda y sus características físicas.

Servicios del hogar.

Variables demográficas.

Salud.

Cuidado de los niños y niñas menores de 5 años.

Educación (personas de 5 años y más).

Fuerza de trabajo (personas de 12 años y más).

Trabajo infantil.

Tenencia y financiación de la vivienda.

Condiciones de vida del hogar.

Tecnologías de información y comunicación.

La encuesta en total tiene más de 700 variables, haciendo casi imposible o por lo menos algo infructuoso el análisis de la totalidad de las variables, así que bajo ciertos criterios estadísticos y sociales que mostraremos más adelante se escogieron las variables sobre las que se iba a trabajar para hacer el análisis.

Lecturas afines y entendimiento sociológico del problema:

El fenómeno de las madres solteras en Colombia es algo que ha venido en aumento en los últimos años, según un artículo publicado por la revista FUCSIA en abril de 2015, nos dice que la tercera parte de las mujeres que son madres en Colombia, están solteras y son jefes del hogar, se da también un dato alarmante: Que el 31% de estas madres cabeza de hogar se encuentran en situación de extrema pobreza y sólo el 8,4% de estas tienen formación profesional. Aunque posiblemente los datos mencionados anteriormente hayan variado un poco debido a que el artículo ya tiene más de tres años con respecto a la fecha actual, se recomienda leerlo. Si bien el objetivo de este trabajo no es analizar la parte sociológica de este asunto, sino puramente la parte estadística, no se puede ser ajeno ya que este es un tema relevante para el progreso del país. Después de esta pequeña reflexión, procedemos a la parte técnica del estudio.

Proceso de filtrado de los datos:

Para la encuesta nacional de calidad de vida 2017, el DANE pone a disposición en su página las siguientes 15 tablas:

Datos de la vivienda.

Servicios del hogar.

Características y composición del hogar.

Salud.

Atención integral de los niños y niñas menores de 5 años.

Educación.

Fuerza de trabajo.

Tecnologías de información y comunicación.

Trabajo infantil.

Tenencia y financiación de la vivienda que ocupa el hogar.

Tenencia y financiación de la vivienda que ocupa el hogar (escritura personas).

Condiciones de vida del hogar y tenencia de bienes (programas).

Condiciones de vida del hogar y tenencia de bienes (programas_2).

Condiciones de vida del hogar y tenencia de bienes.

Seguridad alimentaria.

En la tabla “Características y composición del hogar”, se encuentra una pregunta referente al sexo, con la cual se filtran todas las mujeres; posteriormente se encuentra una pregunta sobre el estado civil, en la que se encuentran las siguientes posibles respuestas: “No está casado(a) y vive en pareja hace menos de dos años”, “No está casado(a) y vive en pareja hace dos años o más”, “Está viudo(a)”, “Está separado(a) o divorciado(a)”, “Está soltero(a)” y “Está casado(a)”, al filtrar las mujeres, se filtran las que estén viudas, separadas o divorciadas o las que estén solteras, después de este filtro, sólo falta verificar que sean madres de alguien que esté en el mismo núcleo familiar. Se elige sólo analizar los casos de madres solteras, que vivan en el mismo hogar con al menos un hijo, esto principalmente porque con la arquitectura de la encuesta no se pueden hallar las madres solteras que no vivan con algún hijo, además consideramos que tampoco sería un caso de interés. De los 40.359 datos iniciales en esta tabla, se halló que se tratan 12.877 casas diferentes, a partir de estos datos, se procede a hacer otro filtrado dónde se garantiza que la madre sea jefe de hogar, ya que así se garantiza la presencia de algunas variables que más adelante se mencionarán. Al finalizar con el filtrado, se halló que en total hay 943 madres solteras de interés, sobre las cuales se procede a hacer un fuerte análisis con respecto a diversas variables que miraremos a continuación. Nota: es importante mencionar que en la base de datos se encontraron algunas preguntas sin ningún valor lógico o null para algunas madres, así que estas se eliminan para que no afecten los resultados.

Las variables o preguntas escogidas para analizar fueron las siguientes:

Cantidad de hogares completos

Estrato

Ingresos mensuales del hogar

Edad

Nivel de satisfacción general con la vida (Esta es la variable respuesta)

Nivel de satisfacción general con los ingresos

Nivel de satisfacción general con la salud

Nivel de satisfacción general con seguridad

Nivel de satisfacción general con el trabajo

Qué tan feliz se sintió ayer

Sintió tranquilidad en la mayor parte del día de ayer

Sintió preocupación en la mayor parte del día de ayer

Sintió tristeza en la mayor parte del día de ayer

En general siente que vale la pena lo que hace con su vida

Estado de salud general

Tiene alguna enfermedad crónica

Edad del primer hijo

Nivel de educación máximo alcanzado

Hactividad que realiza la mayor parte del día

Frecuencia de uso del internet

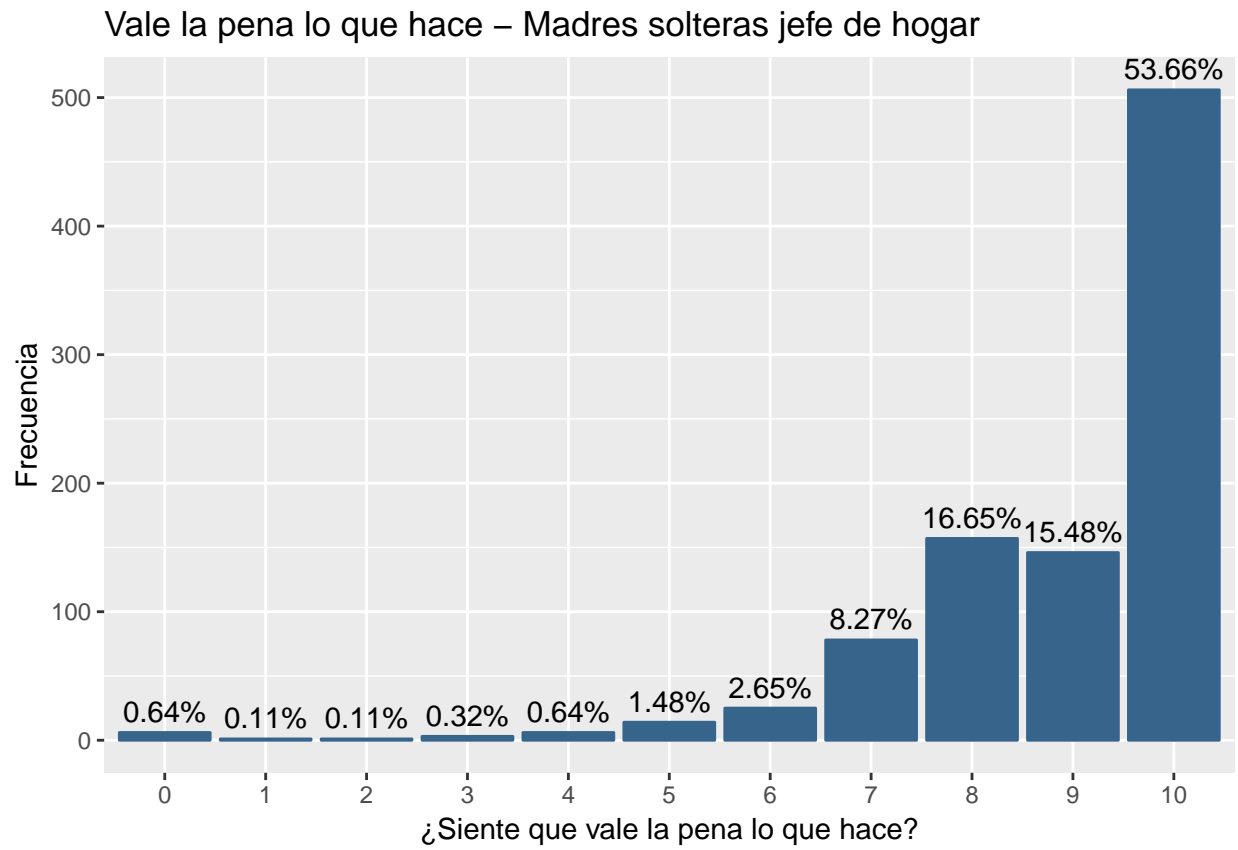
Tenencia de la vivienda ocupada

Condiciones generales de vida

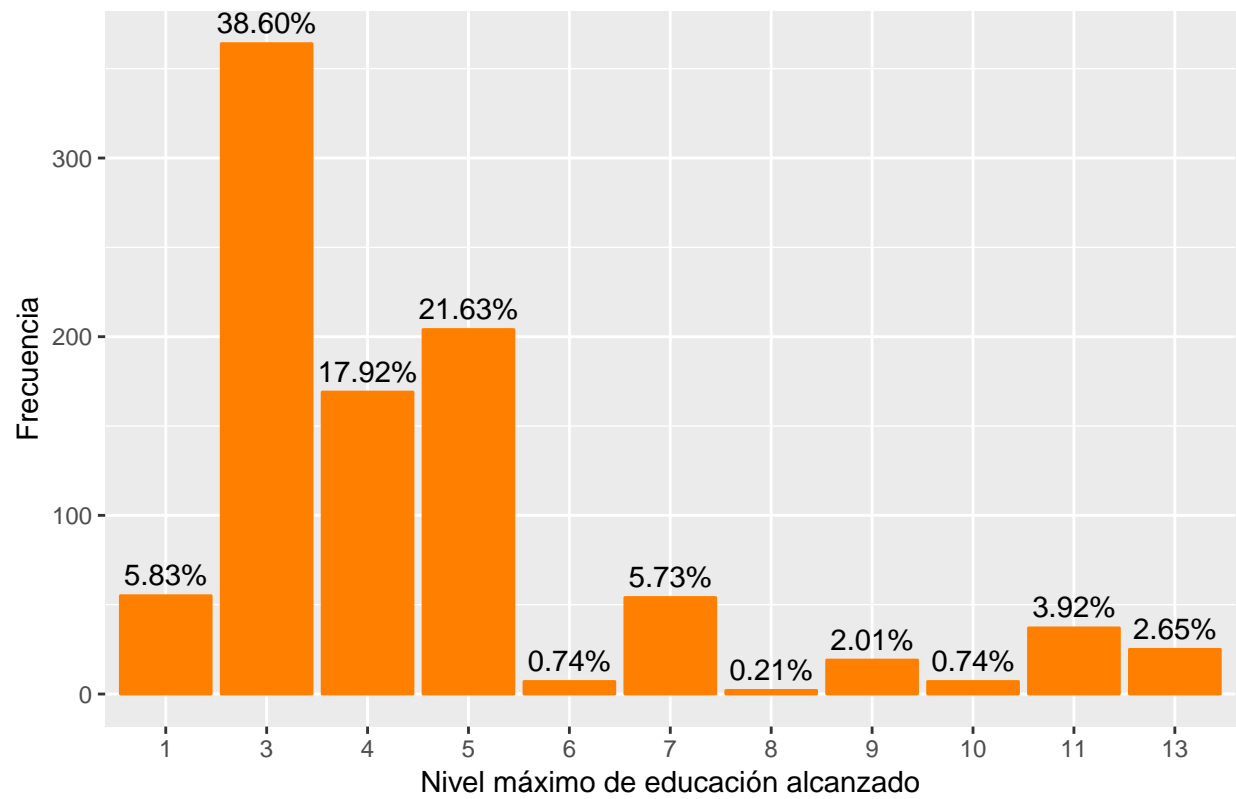
Se considera pobre

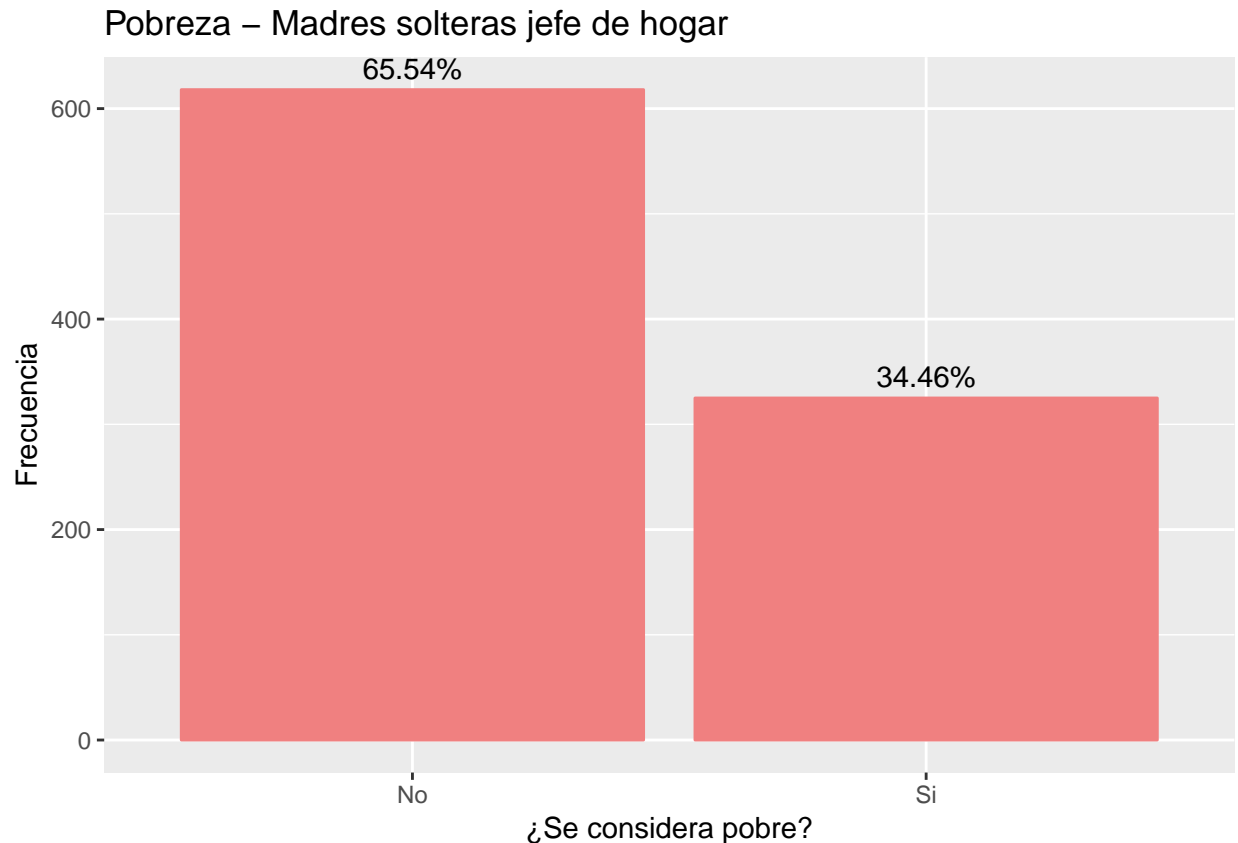
Considera que los ingresos que tiene son suficientes

Además de estas variables, se creó la variable “Edad promedio de los hijos con los que vive”.A continuación veremos un pequeño análisis exploratorio de los datos:



Nivel educación – Madres solteras jefe de hogar





Selección de variables finales para la regresión

Posteriormente se aplicó a estas variables se les aplicó el método de selección hacia atrás y hacia adelante, ambos métodos nos retornaron la misma selección para el modelo predictor:

Nivel de satisfacción general con los ingresos

Nivel de satisfacción general con la salud

Nivel de satisfacción general con seguridad

Nivel de satisfacción general con el trabajo

¿Qué tan feliz se sintió ayer?

En general siente que vale la pena lo que hace con su vida

Condiciones generales de vida

Tenencia de la vivienda ocupada

Edad promedio de los hijos con los que vive

Estrato

Las primeras 6 de estas variables, tienen una escala de 0 a 10 donde 10 es lo que representa “mayor satisfacción” y 0 lo contrario, las otras cuatro variables, tienen unas escalas categóricas diferentes, a las que se les realiza unas simples transformaciones para representarlas como números ordenados donde el mayor número sea el que represente mayor calidad de vida. Después de todo esto, se le aplica una regresión lineal (dada en R por el comando `lm()`) a estas diez variables y a su variable respuesta que es la satisfacción general con la vida, obteniendo así un modelo de predicción con un R^2 de 0.4602 el cual nos da a inferir que el modelo sólo puede explicar un 46% de los datos.

Al ver que el parámetro R^2 no es el mejor, se procede a aplicar algunas transformaciones tanto lineales como logarítmicas:

$$F = \frac{X - \bar{X}}{SD(X)}$$

$$[\text{signo}(F)] \ln(1 - |z|)$$

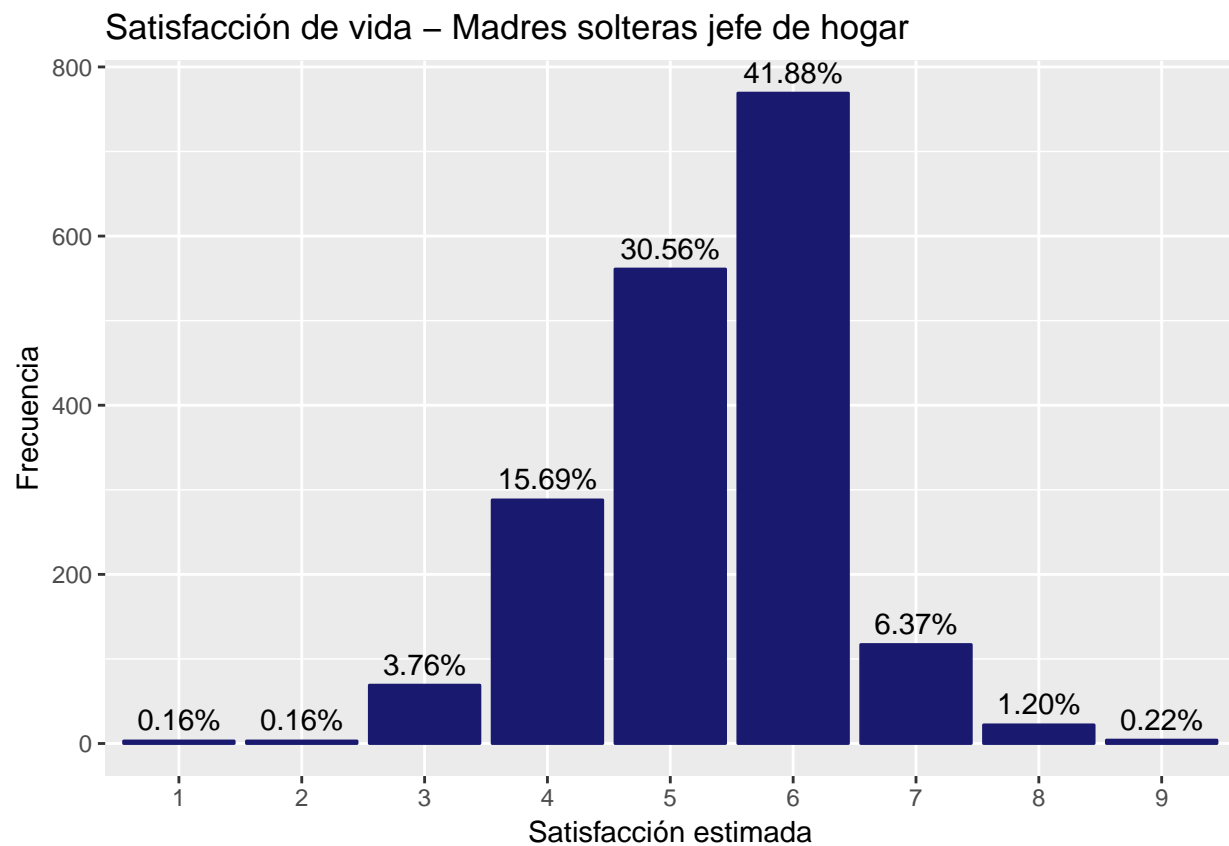
Se le aplica esta transformación en diferentes combinaciones, tales como: A la variable respuesta únicamente, a todas, a algunas regresoras en diferentes combinaciones, etc. Al revisar el `summary()` de R, se puede apreciar que con cada intento, sólo se logra disminuir el R^2 , por lo tanto se descarta inmediatamente las transformaciones y se procede a volver al modelo original.

Luego, se intenta con la siguiente transformación:

$$\ln\left(1 + \frac{(X - \bar{X})}{10}\right)$$

Se sigue un procedimiento similar aplicándolo a distintas combinaciones de variables tanto como a la regresora meramente, pero sólo se logra empeorar el modelo, por lo tanto se descarta.

Prosiguiendo con la búsqueda del mejor modelo posible, procede a realizarse una regresión logística multinomial. Al aplicar este modelo, se obtuvo un missclassification error(Error de el modelo respecto a los datos originales) de el 57%, por lo que se ve que es un error mayor al modelo de regresión lineal clásico inicial, por lo tanto no se descarta el modelo inicial y procedemos a usarlo como el modelo final ya que fue el mejor que se pudo calcular.



```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000  5.000   5.000   5.347  6.000   9.000
## [1] 5.347495
```

[1] 1.021912

[1] 1.044304

	Número	Grado
1	Ninguno	
2	Preescolar	
3	Básica Primaria (1º - 5º)	
4	Básica secundaria (6º-9º)	
5	Media (10º-13º)	
6	Técnico sin título	
7	Técnico con título	
8	Tecnólogo sin título	
9	Tecnólogo con título	
10	Universitario sin título	
11	Universitario con título	
12	Postgrado sin título	
13	Postgrado con título	

Aparentemente se puede ver que los datos se distribuyen normalmente, ya que los datos se están centrado alrededor de la media que es 5.34 aproximadamente con varianza 1.04, por lo tanto se puede entender y ver que la mayoría de datos están cerca a la media la mayoría, más o menos entre 4 y 7.

Aplicación web

Para este trabajo, se nos encomendó realizar una aplicación web, que tuviera dos funcionalidades: La primera es poder seleccionar un hogar y visualizarlo como una red en la que se muestre la estructura familiar, la segunda que debe tener, es que con base en el modelo de predicción explicado anteriormente, la persona que ingrese a la aplicación, al ingresar los datos que esta pide, genere una estimación de la satisfacción. A la aplicación se puede acceder por medio del siguiente enlace.

Video promocional

El vídeo se hizo con la finalidad de vender el producto. Para esto, tuvimos que definir un público objetivo al cual podría serle útil la aplicación y encontrar un problema que sea posible de resolver con nuestro producto. Así, encontramos un estudio en el que se menciona el hecho de que los niños que crecen en un ambiente monoparental, son más propensos a cometer actos delictivos. Bajo esta conclusión, definimos que el público objetivo al que va dirigido el vídeo será el gobierno o entidades que trabajen disminuyendo los índices de delincuencia en el país. De este modo, desarrollamos el vídeo publicitario mostrando nuestro producto como una solución a este problema, teniendo en cuenta que las variables que impactan en la satisfacción de la madre, son variables basadas en el entorno en el que ella se encuentre, y por consiguiente, también impactan a sus hijos. Es decir, una madre que tiene un nivel de satisfacción alto es porque las variables con respecto a su entorno tienen valores altos, lo que se traduce en que ella vive en un buen entorno y por consiguiente sus hijos también (núcleo familiar). Cabe aclarar que el enfoque de lucha contra la criminalidad no es el enfoque de la aplicación sino el enfoque que se le dio al vídeo basado en el público al que se le quiere vender, sin embargo, es posible orientarlo a otro público que pueda serle útil la misma aplicación. El video se puede ver en el siguiente enlace